BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 3: e37078
doi: 10.3897/biss.3.37078

OPEN ACCESS

Conference Abstract

Mobilizing Data from Taxonomic Literature for an Iconic Species (Dinosauria, Theropoda, *Tyrannosaurus rex*)

Jeremy A. Miller[‡,§], Yanell Braumuller[‡,|], Puneet Kishor[§], David Peter Shorthouse[¶], Mariya Dimitrova[#], Guido Sautter[¤], Donat Agosti[§]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ www.plazi.org, Bern, Switzerland
| Leiden University, Leiden, Netherlands
¶ Agriculture & Agri-Food Canada, Ottawa, Canada
# Pensoft Publishers, Sofia, Bulgaria
¤ IPD Böhm, Karlsruhe Institute of Technology, Karlsruhe, Germany

Corresponding author: Jeremy A. Miller (jeremy.miller@naturalis.nl)

Received: 11 Jun 2019 | Published: 13 Jun 2019

Citation: Miller J, Braumuller Y, Kishor P, Shorthouse D, Dimitrova M, Sautter G, Agosti D (2019) Mobilizing Data from Taxonomic Literature for an Iconic Species (Dinosauria, Theropoda, *Tyrannosaurus rex*). Biodiversity Information Science and Standards 3: e37078. https://doi.org/10.3897/biss.3.37078

Abstract

A vast amount of biodiversity data is reported in the primary taxonomic literature. In the past, we have demonstrated the use of semantic enhancement to extract data from taxonomic literature and make it available to a network of databases (Miller et al. 2015). For technical reasons, semantic enhancement of taxonomic literature is most efficient when customized according to the format of a particular journal. This journal-based approach captures and disseminates data on whatever taxa happen to be published therein. But if we want to extract all treatments on a particular taxon of interest, these are likely to be spread across multiple journals. Fortunately, the GoldenGATE Imagine document editor (Sautter 2019) is flexible enough to parse most taxonomic literature.

Tyrannosaurus rex is an iconic dinosaur with broad public appeal, as well as the subject of more than a century of scholarship. The Naturalis Biodiversity Center recently acquired a specimen that has become a major attraction in the public exhibit space. For most species

© Miller J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

on earth, the primary taxonomic literature contains nearly everything that is known about it. Every described species on earth is the subject of one or more taxonomic treatments. A taxon-based approach to semantic enhancement can mobilize all this knowledge using the network of databases and resources that comprise the modern biodiversity informatics infrastructure. When a particular species is of special interest, a taxon-based approach to semantic enhancement can be a powerful tool for scholarship and communication. In light of this, we resolved to semantically enhance all taxonomic treatments on *T. rex*. Our objective was to make these treatments and associated data available for the broad range of stakeholders who might have an interest in this animal, including professional paleontologists, the curious public, and museum exhibits and public communications personnel.

Among the routine parsing and data sharing activities in the Plazi workflow (Agosti and Egloff 2009), taxonomic treatments, as well as cited figures, are deposited in the Biodiversity Literature Repository (BLR), and occurrence records are shared with the Global Biodiversity Information Facility (GBIF). Treatment citations were enhanced with hyperlinks to the cited treatment on TreatmentBank, and specimen citations were linked to their entries on public facing collections databases. We used the OpenBiodiv biodiversity knowledge graph (Senderov et al. 2017) to discover other taxa mentioned together with *T. rex*, and to create a timeline of *T. rex* research to evaluate the impact of individual researchers and specimen repositories to *T. rex* research. We contributed treatment links to WikiData, and queried WikiData to discover identifiers to different platforms holding data about *T. rex*. We used bloodhound-tracker.net to disambiguate human agents, like collectors, identifiers, and authors. We evaluate the adequacy of the fields currently available to extract data from taxonomic treatments, and make recommendations for future standards.

Presenting author

Jeremy A. Miller

Presented at

Biodiversity_Next 2019

References

- Agosti D, Egloff W (2009) Taxonomic information exchange and copyright: the Plazi approach. BMC Research Notes 2: 53. https://doi.org/10.1186/1756-0500-2-53
- Miller JA, Agosti D, Penev L, Sautter G, Georgiev T, Catapano T, Patterson D, King D, Pereira S, Vos RA, Sierra S (2015) Integrating and visualizing primary data from

prospective and legacy taxonomic literature. Biodiversity Data Journal 3: e5063. https://doi.org/10.3897/BDJ.3.e5063

- Sautter G (2019) GoldenGATE Imagine Document Editor. Plazi. URL: http://plazi.org/resources/treatmentbank/goldengate-editor/
- Senderov V, Georgiev T, Agosti D, Catapano T, Sautter G, Tuama ÉÓ, Franz N, Simov K, Stoev P, Penev L (2017) OpenBiodiv: an Implementaion of a Semantic System Running on top of the Biodiversity Knowledge Graph. Proceedings of TDWG 1: e20084. https://doi.org/10.3897/tdwgproceedings.1.20084